AD

TECHNICAL REPORT

TR 76-40FSL

JOB KNOWLEDGE TESTING AS AN INTERMEDIATE CRITERION OF PERFORMANCE: A STUDY OF US NAVY COOKS

Approved for public release;
distribution unlimited.

JUNE 1976

UNITED STATES ARMY
NATICK RESEARCH and DEVELOPMENT COMMAND
NATICK, MASSACHUSETTS 01760

Food Sciences Laboratory

Approved for public release; distribution unlimited.

Citation of trade names in this report does not constitute an official indorsement or approval of the use of such items.

Destroy this report when no longer needed. Do not return it to the originator.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR 76-40 FSL
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): JOB KNOWLEDGE TESTING AS AN INTERMEDIATE CRITERION OF PERFORMANCE: A STUDY OF US NAVY COOKS

5. TYPE OF REPORT & PERIOD COVERED:

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Earl S. Stein, Lawrence E. Symington
Herbert L. Meiselman

8. CONTRACT OR GRANT NUMBER(s):

9. PERFORMING ORGANIZATION NAME AND ADDRESS: U.S. Army Natick Research and Development Command, Natick, MA 01760
DRXNM-YBH

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 6.2
1J662713AJ45, Task 03

11. CONTROLLING OFFICE NAME AND ADDRESS: Same as above

12. REPORT DATE: June 1976

13. NUMBER OF PAGES: 40

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release;
distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This work was performed in response to a requirement of the DOD RDT&E program, USN 5-1, to the Operations Research and Systems Office, U.S. Army Natick Research and Development Command

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Food Service
Training Evaluation
Performance
Criterion-referenced test
Performance Evaluation
Skill Evaluation
Job Knowledge
Skills
Training
Job Performance
Job Training

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A job knowledge test was constructed using items from the Navy correspondence course program. The test was administered to 355 Navy Food Service workers from the Atlantic and Pacific Fleets. The Job Knowledge Test was internally reliable and was positively related to supervisors' ratings. Test performance was strongly related to paygrade, with those in the higher grades doing better. Graduates of the Basic Food Service Course ("A" School) performed better than non-graduates.

JOB KNOWLEDGE TESTING AS AN INTERMEDIATE CRITERION OF PERFORMANCE: A STUDY OF US NAVY COOKS

Earl S. Stein

Lawrence E. Symington

Herbert L. Meiselman

Food Sciences Laboratory
U.S. Army Natick Research & Development Command
Natick, MA 01760

PREFACE

The authors wish to express their appreciation to the U.S. Navy personnel who made this report possible, in particular LCDR Robert Helmuth, the U.S. Navy Representative on the Joint Technical Staff, DOD Food RDT&E Program, who coordinated the testing sessions. We would also like to thank Mr. Theodore Mattus of the Operations Research Systems Analysis Office, U.S. Army Natick Research and Development Command who assisted in the administration of the Job Knowledge Test at the various Navy activities.

Table of Contents

LIST OF TABLES

LIST OF TABLES (CONT.)

APPENDIX

SUMMARY

The Navy Mess Management Specialty (MS) involves many different kinds of tasks and represents a complex job system from the viewpoint of measurement.

There are three types of criteria for such measurements: immediate, conducted during training; intermediate, accomplished between training and actual job performance; and ultimate, measures taken during actual job performance. Immediate measures are similar to school tests and generally are the most distant from the job. Ultimate criteria are difficult to define and expensive to measure. These are some of the reasons why a job knowledge test (JKT) was selected as an intermediate criterion to help evaluate Navy mess specialty training.

The test was constructed using Navy correspondence course materials, initially pilot tested, and then administered in final form to 355 Navy MS (pay grades E-1 to E-6) from the Atlantic and Pacific fleets.

The test demonstrated an internal reliability of 0.85. It did not correlate with Navy entrance test scores or most efficiency report scales. Test results were, however, significantly related to seven performance oriented scales that were filled out on most sampled individuals by their supervisors.

Former Stewards tended to score higher on the test than former Commissarymen, at paygrades other than E-6, with the most pronounced difference occurring at the E-5 level. It was hypothesized that increased experience may counteract earlier differences in knowledge.

Paygrade was directly related to test scores; it could generally be concluded that as grade increased test performance increased also. Watch captains did not perform better than their subordinates, and, therefore, they must provide leadership based on skills other than those which contribute to higher test scores.

"A" School graduates demonstrated superior JKT performance over non-graduates at all grade levels tested except for E-6. In grades E-1 through E-3 formal instruction and correspondence course training provided superior performance compared to that of personnel who had no training or only correspondence training.

When all paygrades were pooled the general conclusion to be drawn from JKT scores was that any structured form of training, including correspondence training, was superior to no training. This finding implies a necessity for continuation of formal programs.

Introduction

Training evaluation is necessary for establishing the status of current training and determining how it could be improved. It is also useful in "determining if graduates meet job performance requirements in the field" (Siegel, Bergman, Federman & Sellman, 1972). Training evaluation is a process of making decisions about the value of a training program as a whole or in part. Siegel et al. (1972) note that training studies are commonly either "internal", conducted in the training situation, or "external", conducted in a field situation. It is apparent that systematic evaluation of training is superior to the acceptance of programs at face value or on the basis of opinion statements from responsible individuals (Tiffin, 1952).

The U.S. Navy Mess Management Specialist rate (MS) involves many heterogeneous tasks which are difficult to measure directly. In order to evaluate the training of Navy MS's, one must enter somewhere in the chain of events which includes the training itself and progress through actual performance of the learned skills in the fleet.

At the center of this task is what is called the criterion problem. It requires very specific definitions of what is to be measured and what the practical limitations are. Measures which are taken before training are usually referred to as aptitude scores, while those which are taken during training are called achievement scores. Performance knowledge testing done sometime after training, when the forces of selective retention have had an opportunity to work, can be called job knowledge testing. Criterion measures which are essential for training evaluation may be divided into three categories: immediate, intermediate, and ultimate. (Thorndike, 1949; Foley, 1974). Immediate measures are generally those which

---

Siegel, A.I., Bergman, B.A., Federman, P., Sellman, W.S., Some techniques for the evaluation of Technical Training Courses and student. (AFHRL-TR-72-15) Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, Feb. 1972.

Tiffin, J. Industrial psychology. New York: Prentice-Hall, 1952.

Thorndike, R.L. Personnel selection. New York: Wiley, 1949.

Foley, J.P. Evaluating maintenance performance: an analysis. (AFHRL-TR-74-57(1)), Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, Oct. 1974.

first become available either during or immediately after training completion. The ultimate criterion is the final goal of training. It requires quantification and data collection on job performance in the field. Between the immediate and ultimate extremes on the criteria continuum lie numerous measures which could be classified as intermediate criteria. Among these is job knowledge testing.

Job knowledge testing traces its history to the pre-World War II techniques of oral trade tests which were used as simple employment screening devices to determine if an individual had the training and experience he claimed (Lawshe, 1948).

Job proficiency can be measured in many ways. Measuring job knowledge is based on the premise that underlying skills and knowledge give the individual the capability to do his job (Gagne & Fleishman, 1959). Measuring actual performance as an ultimate criterion requires fewer assumptions but is generally more expensive and very difficult to do. As the tasks in a given work role increase in number and complexity (as they do in the Navy Mess Management Specialists rate) it becomes very difficult to find effective ultimate criteria (Vroom, 1964). Job knowledge testing represents an abstraction from actual performance, and the effect of job knowledge on work can be qualified by individual motivation. Recognizing the liabilities of the job knowledge testing techniques, there are specific advantages. The technique is similar in concept to the promotion testing systems currently in use in all the armed forces. It is superior to summary judgments of an individual which require a rater to comment on a subordinate's performance over a long time span. Problems of interrater reliability and personal bias are avoided, and a permanent test record is left which can be restored if necessary (Thorndike, 1949).

The evaluation of training in the Navy Mess Management Specialty rate is the purpose of this project. Barrett (1966) noted that there

---

Lawshe, C.H. Principles of personnel testing, New York: Mcbran, 1948.

Gagne, R.M., and Fleishman, E.A. Psychology and human performance. New York: Holt, 1957.

Vroom, V.H. Work and motivation. New York: Wiley, 1964.

Barrett, R.S. Performance rating. Chicago: Science Research, 1966.

are various ways you can examine an individual, the process, or the product. Characteristics of the individual and some observations concerning his work (the process) are handled by some sort of efficiency report which generally suffers from many of the problems of global rating, where the individual is rated as good or bad on general criteria which do not specify the behavior that should be observed. The product generally has not been considered.

Our general approach to the behavioral aspects of the Navy Mess Management Specialist training problem was to select the job knowledge test option as a first step in a long term program of training evaluation. A concurrent effort of using general performance ratings was also selected. These were chosen after surveying subject matter experts in the Navy Mess Management Specialst rates. Using such experts was emphasized by Thorndike (1949) and Barrett (1966). The ratings were designed and employed as data distinct from Navy efficiency reports. We guaranteed confidentiality from the Navy personnel system.

There are many questions to be asked. Assuming that the Job Knowledge Test is a successful measure, will it discriminate between individuals on the basis of training received? Do various types of training enhance job knowledge in different ways? Merenda (1958) found that formal school training was superior to on-the-job training (OJT) in various Naval occupations. Would the job knowledge test reflect similar differences for current Navy MS's? The following report deals with these questions.

Merenda, P. F. The relative effectiveness of formal school and on the job methods of training apprentices in naval occupations. Personnel psychology, 1958, 11, 379-389.

Method

Test Construction

Constructing a test in a specialized area of knowledge requires either a thorough acquaintance with the area or having support from those who do. It was initially decided to acquire test items (questions) and item statistics from the Navy promotion test system. This would have permitted item selection on the basis of already established data. However, due to such factors as test security precautions, such information was not available, and alternatives had to be sought.

The Navy provided rate training manuals and correspondence course questions including the following titles: Commissaryman 3+2, Steward 3+2, Commissaryman 1+C, and Steward 1+C. Although Commissaryman (CS) and Steward (SD) rates have now been merged into the Mess Management Specialists rate (MS), it was felt that the basic knowledge covered by these manuals would not change appreciably.

Item selection for the pilot test was of major importance. Four choice items were selected by one author who read the rate training manuals and selected items which seemed to provide coverage of the material and would be of average difficulty. It was decided to construct one test for those in pay grades E-1 through E-6, who serve as cooks and watch captains.

During the early item selection phase, it was believed that approximately 100 items would be required. We wanted the test adequately long to improve the possibility of reliability since time limitations would prohibit a test-retest or alternate form reliability assessment. Also, high internal consistency reliability (i.e., split half or Kuder-Richardson) was not a

---

Bureau of Naval Personnel. Commissaryman 3+2 rate training manual (NAVPERS 10279-E) U.S. Government Printing Office Washington; (1971a).

Bureau of Naval Personnel. Steward 3+2 rate training manual (NAVPERS 10694-D). U.S. Government Printing Office: Washington, (1971b).

Naval Training Command. Commissaryman 1+C rate training manual (NAVTRA 10280-F). U.S. Government Printing Office: Washington, (1973).

Naval Training Command. Steward 1+C rate training manual (NAVTRA 10695-D). U.S. Government Printing Office: Washington, (1972).

primary goal because of the heterogeneous nature of the material. Initial item selection produced a pilot instrument with 139 items. Since these items had come from the four rate-training manuals, they were distributed in the printed test so that each quarter of the test contained an even distribution of items from each manual. No matter what a testee's background, and irrespective of whether he completed the test, he was exposed to a balanced variety of items. These are primarily technical considerations noted here for the use of other behavioral professionals and the lay reader may ignore them if he chooses.

Pilot Test

A pilot test with no time limit was conducted at the Newport Naval Education and Training Center with 24 participants; both former Stewards and Commissarymen. Table 1 indicates that there was a sizable score range, which is a desirable outcome. Item difficulty statistics were also favorable.

Item difficulty is the number of individuals answering an item correctly, divided by the total number of people taking the test. This provides scores from zero (0) to one (1.0) where everyone gets a given item correct. The item difficulty range in this pilot test was in fact from zero to one. Since an item difficulty of 0.5 provides the best opportunity to separate individuals (Guilford, 1954, p. 365), it was desirable to move the difficulty range closer to this ideal. Items with difficulty scores above 0.80 (80% of the testees answered correctly) and below 0.20 were eliminated. Twenty seven items fell into these groups leaving a final test of 112 items.

During the initial pilot, completion time was recorded. A number of individuals took more than one hour to finish. It was felt that a test of more than one hour length, when administered in conjunction with attitude surveys, would be too much of an operational imposition on ship commanders. Fortunately the length was already reduced on the basis of item difficulty, and further cuts were not considered. However, recognizing that the final instrument as administered (with a 1 hour time limit out of practical necessity) would be a compromise between a power and a speed test[1], it was hoped that the speed component would

---

Guilford, J.P. Psychometric methods McGraw: New York, 1954.

[1] Power tests - imposes no time limit, while speeded tests do.

Table 1

Initial Pilot of the Job Knowledge Test

Category	Number	Mean Score (Number Correct)	Mean Completion time (min)	Range Low	Range High
Commissarymen	10	66.5	51.6	40	85
Stewards	14	71.7	84	48	92

Table 2

Number of Personnel Tested

	East	West
Commissarymen:	106	93
Stewards:	67	60
Unspecified:	15	14
Total	188	167

not lead to the contamination of results by such factors as reading speed and verbal intelligence.

Administration

Administration Conditions

It has been noted that the Job Knowledge Test was limited to one hour of administration time. It was given in conjunction with an untimed opinion survey which took most participants about 30 to 40 minutes to complete. In order to control for possible fatigue and attitudinal effects, the administrative order was counterbalanced so that approximately half the sample received the JKT first and the other half received the survey first. No matter which order was to be used the participants received the two instruments at the beginning of the session. The administrator gave a five to seven minute briefing covering both. In addition, detailed instructions for the JKT were provided on the first page of the test. One administrator was present at all times to answer questions, except for those concerning item content.

The test and the survey were administered aboard ship and/or in shore classrooms depending on the easiest arrangement for the various supply and food service officers. The classroom settings were well ventilated, well lighted and quiet. Shipboard conditions were more variable. For the most part administration was conducted in messing areas where lighting and ventilation were adequate. While the noise level and interruptions varied from ship to ship, the personnel seemed to be adapted to the conditions and simply kept working. Most work on the west coast was aboard ship in contrast to the east coast where it was primarily in a classroom.

Sampling

The Job Knowledge Test (JKT) was administered to 355 food service workers on the east and west coasts (Table 2). The "Unspecified" category includes individuals who have joined the Navy since the CS-SD merger and also some who did not state whether they had been CS's or SD's under the old system. Ships were sampled by size class. The selection of the specific ships was left to the Navy and the selection of individual participants remained the prerogative of the various supply and/or food service officers.

Technical Results and Discussion

Completion Rate

It has been noted that the Job Knowledge Test was speeded to a certain extent because of administrative necessity. One index of the degree of speeding is the number of items attempted. Although there were 112 items, not everyone was able to complete the instrument. Table 3, represents the simplified summary of the percentage of individuals completing a certain number of items. The completion statistics indicate that a sizable portion of the sample (74% East and 50% West) were able to complete the entire test. There are differences in completion rate between the Atlantic and Pacific fleets, which may be attributable to differences in administration conditions and/or to other unspecified factors. Since, as will be shown later, there was no difference between the fleets in JKT performance, respondents in the Pacific fleet may have had a slightly higher correct percentage of items attempted.

Item Analysis

Because of the speeded component, it was decided to do an item analysis on only the first 80 items and to make the assumption that they will be representative of the entire test.* Approximately 95 percent of those taking the test completed these items. Data from 355 participants were computer analyzed using the BMD factor analysis program (using orthoganal analysis with verimax rotation). The average item difficulty was 0.45 which was close to the goal of 0.50 generally considered ideal for separating individuals on the entire test (Guilford, 1954).

The analysis generated ten rotated factors which could account for only about one-fifth of the total variability. What this indicated was that many of the items were contributing to the total score in their own unique way, and that their contribution was not redundant with other items. This may indicate that the sampled items covered a broad spectrum of the knowledge area.

The JKT and Navy Entrance Exams

The JKT was constructed to measure specific job knowledge and was in no way intended to be a general aptitude scale. Because of the reading requirement and the speeded administration, it

* NOTE: The validity of this assumption was supported by a part-whole (80 items vs 112 items) correlation of .91.

Table 3

Percentage of Individuals Completing Items

ATLANTIC FLEET

% of Sample Completing Items	No. of Items Completed
100	54
95	86
90	92
85	98
80	106
75	111
74	112

Mean No. of Items Completed 107.5

PACIFIC FLEET

% of Sample Completing Items	No. of Items Completed
100	55
95	77
90	82
85	89
80	91
75	95
70	97
65	102
60	107
55	111
50	112

Mean No. of Items Completed 102.9

was hypothesized that some elements of what might be referred to as "paper and pencil or verbal intelligence" could be influencing the overall JKT scores of individuals. In order to determine if this was a problem, it was decided to compare JKT scores with scores from the Navy General Classification Battery supplied by the Bureau of Naval Personnel. This was possible for 271 individuals or 76% of our sample. It was believed that these entrance tests were psychometrically sound and contained a reasonable index of intelligence as it is commonly conceptualized. Table 4 indicates that JKT performance was not related to general

TABLE 4

Correlations Between JKT and Navy Aptitude Scores

Navy Aptitude Test	Correlation
General Classification Test	-.048
Arithmetic Test	.044
Mechanical Test	.111
Clerical Aptitude Test	.043

aptitude as measured by the Navy entrance tests. These results reinforce the likelihood that the JKT is actually measuring job specific knowledge, not general intelligence.

Test Validation

Test validation demonstrates one of the oldest problems in applied industrial psychology. This is the criterion problem or the definition of what we want to predict. The Job Knowledge Test was designed as a research instrument to determine if there were differences in basic knowledge across various groups of Navy MS's. However, without any validity information one might rightfully question the meaning of such differences.

We decided to examine the relationship of JKT scores to two types of performance related criteria. One was the standard Navy efficiency report. Table 5 shows the matching of JKT scores with efficiency scores for 81 individuals on nine scales and for 191 on four scales. This discrepancy resulted from missing data on the magnetic tapes supplied by the Navy. The first ten scales demonstrate very low nonsignificant correlations with the JKT.

Table 5

Correlations Between JKT and Efficiency Report Scores

Efficiency Scale	Correlation
Performance	.096
Appearance	.056
Cooperation	.153
Reliability	.116
Conduct	.076
Resourcefulness	.110
Leadership-Direction	.134
Leadership-Counseling	.110
Overall Evaluation	.109
Recruiter Potential	.166
Instructor Potential	.206*
Career Coun eling	.247**
Independent Duty	.271**

* P <.05

** P < .01

The last three scales, although presenting low correlations, are related to JKT performance in a positive and significant manner. These three scores represent the rater's prediction of an individual's potential rather than his evaluation of actual performance. The general trend of the Navy Efficiency scores is toward the military aspects of an individual's job. The Job Knowledge Test was oriented specifically on non-military aspects of Navy cooking.

The low correlations of the efficiency scores and the JKT were anticipated, and for that reason our own performance ratings were developed. Members of several Food Management Teams (senior food service trainers) were asked to write down critical categories of behavior which they felt were important and observable. These categories were edited and the following six scales were constructed: cooking ability, supervisory skills, motivation, personal hygiene, job knowledge, and attitude toward other workers. On each scale leading MS's were asked to rate their men individually on a seven point scale from 1 (very bad) to 7 (very good). These ratings were accomplished at the same time that the ships' MS's were taking the surveys and

Job Knowledge Test. This is in contrast to the efficiency scores which could have been done weeks, months or years prior to JKT administration.

The correlations between these Performance Scales and JKT scores were better than the efficiency scores. All validity coefficients (correlations) were significant, different from zero ($P < .01$) and are reported in Table 6. The supervisors were repeatedly assured that their ratings would not be communicated to any agency which could influence the promotion or retention of their men. They apparently believed this and generated ratings which showed variability between individuals and across scales within individuals.

Table 6

JKT Validity Using Performance Ratings

Rating	Correlation with JKT	N
Cooking Ability	.384	175
Supervisory Skills	.381	163
Motivation	.365	191
Personal Hygiene	.315	191
Job Knowledge	.393	188
Attitude Toward Other Workers	.314	191

These differences enhanced the possibility of relationships with the JKT scores. It is interesting to note that the performance rating which provided the highest correlation with the JKT was "job knowledge".

Test Reliability

No matter what a test purports to measure, it must do it consistently to be of any value. Due to the fact that neither two administrations of the JKT nor the construction of a parallel form were possible, reliability estimation was limited to measures of internal consistency. Kuder-Richardson reliability (KR formula 20) was computed for the first 80 items in the test. Only the first 80 were employed due to the decline in completion rate toward the end of the test. The reliability coefficient was 0.85. Since the test was made up of homogeneous items an estimate was made of the potential reliability of the entire 112 item test where unlimited time would have been available for administration. Using the Spearman-Brown Prophesy Formula this estimated reliability was 0.89. Coefficients such as these are very acceptable in a research test which measures heterogeneous skills. The JKT was not designed as a marketable

instrument and was also not geared for a homogenous content area which would necessitate higher internal consistency.

Overall Results and Discussion

What has been considered up to this point are data concerned with the nature of the Job Knowledge Test and its relationship with other measures. The major purpose of the test was, however, to determine job knowledge in a very specific sense relative to certain demographic variables: paygrade, education, the distinction between Commissarymen (CS) and Stewards (SD), and the Atlantic vs. the Pacific fleet. The remainder of this report will concentrate on these variables.

An astute reader who attempts to sum up various subsample sizes to reach the total sample may become quite frustrated. First, the totals include individuals who classified themselves as neither Commissarymen (CS) or Stewards (SD); these include individuals who joined the Mess Management specialty after the merger. Adding the numbers for the CS's and SD's will not give you the grand total. In addition, whenever the total sample was broken down into any categories (such as pay grade or educational grouping) if there was any doubt as to an individual's membership in a class, he would be dropped from that particular analysis.

Atlantic vs. Pacific fleets

Mean JKT scores from the Atlantic and Pacific fleets were compared and showed no statistically significant difference ($t=.89$ $P > .05$) (table 7). Although this was the situation for

Table 7

The Overall Sample (East vs West)

	ATLANTIC	PACIFIC
Mean	47.6	46.28
Number	188	167

the total sample, some data will be analyzed separately for east and west in case there were subtle differences in specific subsamples.

Commissaryman (CS) - Steward (SD)

Several Navy food service personnel (both officer and enlisted) have expressed a great deal of interest in the CS-SD merger. Personnel in the fleets assumed that this survey/testing was to assess the results of the merger. While such an examination was not the intention of this project, the possibility of differences between former CS and SD personnel in terms of job knowledge was addressed. As can be seen in Table 8, former SD's tended to score significantly higher than their former CS counterparts in both fleets and in the pooled data. This is supported by the tests in Table 9.

Table 8

Average JKT Scores for Stewards & Commissarymen

	East		West		Pooled	
	SD	CS	SD	CS	SD	CS
Number	67	106	59	91	126	197
Mean Score	52.22	44.76	50.33	44.23	51.34	44.5

Table 9

t Test Results for Steward - Commissaryman Comparisons

	East	West	Pooled
t	3.796**	2.53**	4.47**
eta^2	.078	.041	.058

** P <.01

These eta^2's result from a technique referred to as strength of association tests. They indicate the amount of variability in JKT scores that can be accounted for by the independent variable - in this case the CS-SD comparison. Any eta^2 below .10 (or 10%) of the variability should not be viewed as a strong difference

(Linton & Gallo, 1975). What this means is that although the stewards tend to score somewhat higher, they do not score that much higher as a group.

Employing data with the Atlantic and Pacific combined, an examination of varying CS-SD differences at each grade level was accomplished. A paygrade by steward-commissary analysis of variance (ANOVA) produced significant steward-commissary and paygrade main effects ($P < .01$), but no interaction. The steward-commissary differences were analyzed at each grade level (Table 10). The strength of the steward-commissary difference varied across pay

Table 10

Stewards - Commissarymen Comparisons by Paygrade

Paygrade	t ratio	eta^2
E (1-3)	1.75*	.009
E-4	4.64**	.0635
E-5	6.42**	.1151
E-6	.78	

* $P < .05$
** $P < .01$

grades. Those in paygrade E-5 were considerably different. The E-5 SD performed quite a bit better than his counterpart. The E-4 SD also performed to a superior degree but not to the same extent. At the E-6 level there was no difference at all.

The primary differences are in the middle of the grade range of those sampled. Most of the E(1-3) have entered the Navy during the merger, and perhaps have not been in the system long enough to differentiate their abilities. In contrast, the E-6 (MS-1, Petty Officer 1st class) personnel have had lengthy on-the-job experience and training which could produce a broader knowledge base.

Linton, M. and Gallo, P.S. The practical statistician: simplified handbook of statistics. Monterey: Brooks, 1975.

Pay Grade and JKT Performance.

The relationship between paygrade and JKT score is demonstrated by rank correlations between the variables (see Table 11). ANOVAS were run on paygrades E 2-6 to further examine this relationship.

Table 11

Spearman - Rho Correlations Between Paygrade and JKT Scores

	Stewards	Commissarymen	Pooled
Atlantic	.975	.875	1.0
Pacific	1.0	1.0	1.0

There were not enough E-1's (N = 3) to consider in the analysis. The F values provided by the ANOVAS were significant in all cases at the P <.01 level. See Table 12 (and A-1 through A-8, in Appendix A

Table 12

Table of Differences Between Means - The Results of Analysis of Variance and the Newman - Keuls Procedure
All Personnel - East and West Pooled

Grade		2	3	4	5	6
GRAND TOTAL:	Means	37.65	39.98	50.08	53.64	61.60
ANOVA F=43.10**	2	37.65	2.33	12.43*	15.99*	23.95*
*P <.05 **P <.01	3	39.98		10.1*	13.66*	21.62*
	4	50.08			3.56	11.52*
	5	53.64				7.96*
	6	61.60				

Grades sharing common underline do not differ: Schematic Presentation of Differences 23 45 6

for details). Table 13 presents strength of association measures for all of the ANOVAS. The values range from 0.26 to 0.36 which

Table 13

Strength of Association Measures for the ANOVA Designs on Paygrade W^2 ($OMEGA^2$)[1]

	Stewards	Commissarymen	Stewards & Commissarymen Pooled
East	.36	.34	.34
West	.30	.26	.30
E-W Pooled	.33	.30	.32

[1] Each coefficient indicates the proportion of variance accounted for on the basis of the independent variable of paygrade. By current standards any measure over .1 may be looked at as being a real effect.

are large enough to conclude that paygrade is an important (of practical significance) variable in relation to JKT performance.

Paygrade is important to job knowledge test performance for all personnel. This alone is a very relevant finding. It was suspected that there might be some serious rank reversals in JKT performance because of recent school graduation or correspondence course completion by individuals in the lower pay grades. This did not occur. To the extent that the JKT measures job knowledge, then the combination of education and experience available to Navy mess specialists is working. On the whole an E-6 (CPO 1st class) knows more than an E-2 seaman although many of the latter are recent school graduates.*

The ANOVA demonstrates that there were differences in test performance between paygrades E-2 and E-6. To find out exactly where the differences were, a Newman-Keuls procedure was employed. This is accomplished by ordering the mean scores in a table such as those in Table 12. Then difference scores are computed between the means and compared to a statistical table. Those scores that are significantly different are indicated with an asterisk. Below the table is a schematic presentation of differences indicating that pay grades 2 and 3 are not significantly different from

*The possibility does exist that personnel in the higher paygrades, E 4-6, had previous experience with some of the JKT questions in their correspondence courses.

each other; neither are 4 and 5. However, paygrades 2 and 3 are different from 4 and 5 and all are different from paygrade 6. This can be confirmed by looking at the asterisks in the table itself.

Watch Captains.

In the overall sample, 78 watch/galley captains were tested. It could be hypothesized that due to their leadership status their test performance might be different from that of cooks in the same paygrades. Table 14 presents the mean scores for these groups.

Table 14

Average JKT Scores as a Function of Leadership Status and Paygrade (Sample no. in parenthesis)

	Paygrade			
Leadership Status	3	4	5	6
Watch Captains	41.8	48.76	51.35	61.34
	(5)	(21)	(20)	(32)
Cooks	39.89	50.57	54.82	62.36
	(94)	(65)	(39)	(11)

A (2 x 3) analysis of variance examined leadership status along with paygrades 4, 5, and 6. Paygrade 3 was not included because of the small number of watch captains involved. There was no effect of leadership status ($F = .379$) and no interaction ($F = .001$). However, the paygrade variable, as usual, was significant ($F(2,72) = 4.11$, $P < .05$). These results imply that watch captains do not have more academic knowledge than their subordinates in the same paygrades. This suggests that whatever leadership skills they possess are based on something other than superior knowledge in a paper and pencil sense. Possibilities would include superior performance skills based on experience and/or a highly developed ability in interpersonal relations. Should both of these be lacking, one choice left for a watch captain would be to fall back on the legitimized power inherent in his position. This could place him in a difficult situation and enhance subordinate dissatisfaction.

Education.

The effects of inservice training are both the most important and the most elusive aspects of this project. Training takes

many forms including, but not limited to, formal A & C schools, correspondence courses and on-the-job training (OJT). "OJT" has to be considered part of the individuals overall experience since it often has no defined onset of completion. Also individuals may differ when asked if they have had OJT in a particular area, because in many cases it can be so subtle that the individual is not aware that he is being trained.

Formal Training.

We examined the effects of formal training on JKT performance by paygrade since this variable seems to overshadow the effect of all others. The mean JKT scores of individuals in various educational classifications demonstrate the usual paygrade effect (Table 15).

Table 15

Average JKT Scores as a Function of Formal Training

	No School		A only		C only	
Grade	Mean	No	Mean	No	Mean	No
1-3	34.7	47	41.2	120		
4	46.4	24	52.5	56	47.6	3
5	46.9	21	59.3	26	53.6	7
6	63.9	16	64.8	15	58.8	12

What is notable is the number of MS's who have not had any formal schooling, especially those in grades E 4-6. A sizable number was expected in grades E 1-3 because these people in many instances are strikers who entered food service from other specialties. Also there are very few C school graduates relative to the overall sample size. This may be partially a function of the fact that data was collected only on E 1-6, and C graduates tend to be in the higher grades. The fact that the C graduates average score tends to drop below that for A graduates should not be taken very seriously because of the small numbers involved and the possibility that the sample was somewhat biased.

Table 16

Results of t tests between No School and A School Graduates

Grade	t	df	Strength of Association
1-3	4.336**	165	.102
4	2.00*	78	.049
5	3.54**	46	.214
6	.18	29	NA

** P < .01
* P < .05

Table 16 shows the effect of A School on JKT performance as compared to no school at all. The superiority of A graduates is somewhat questionable in grade E-4 but quite pronounced in the others. In contrast, graduating from A School had no effect on respondent's scores in paygrade E-6. The on-the-job experience gained by those at the E-6 level no doubt has a leveling effect, and formal schooling becomes less important.

Formal vs. Correspondence Training.

Table 17 compares formal schools and correspondance training.

Table 17

Average JKT Scores As A Function of Different Types of Training

	No School or Correspondence		Correspondence Only		Formal School[1] No Correspond.		Both Formal & Correspond.	
Grade	Mean	No.	Mean	No.	Mean	No.	Mean	No.
1-3	34.27	26	35.3	21	40.9	64	40.8	59
4	47.5	4	46.2	20	47.9	11	53.1	49
5	45.6	3	46.9	18	57.25	4	57.6	33
6			63.5	16	57	3	63.03	29

[1] A and/or C Schools.

For the purposes of this analysis, correspondence training was defined as having taken one or more correspondence courses, and formal school was defined as having completed A and/or C school. (C school could not be treated separately because of the small N's involved). The most informative method of analyzing the data in Table 17 would have been a grade by schools (4 x 4) analysis of variance. Since the cell sizes were so unbalanced this was not deemed possible. We decided to break up the design and ran two ANOVA's, one across schools at the pay grade 1-3 category and another across schools with all pay grades pooled. Selective t tests were also accomplished at the E-4 and E-5 levels.

The ANOVA across schools at grade E (1-3) indicated a significant training effect on JKT scores. ($F(3,166) = 5.646^{**}$, $P < .01$). The Newman-Keuls procedure was used to determine where the differences occurred. Table 18 provides the differences between the ordered pairs of means for the various training categories.

Table 18

Differences Between Mean JKT Scores on the Basis of Different Training Categories for Paygrade E 1-3 (The Newman-Keuls Procedure)

	No School	Correspondence Only	Both Formal and Corresp.	Formal School Only
Means	34.269	35.33	40.78	40.9
No School 34.269		1.061	6.511*	6.637*
Corresp. only 35.33			5.45*	5.576*
Both Frml. & Corres. 40.78				.126
Frml Sch. Only 40.9				

$p < .05$

Schematic Presentation of Significant Differences:

No School	Correspondence Only	Formal School & Correspondence	Formal School Only

Categories which share common underlining are not significantly different.

Since the JKT was constructed using items selected from the correspondence course system, one might expect that those who have completed correspondence programs would perform better on the JKT than those who had not. However, for those in grades E (1-3) the correspondence courses provided no advantage over no school at all and no advantage over any formal training. Formal schools led to significantly better scores than no training and correspondence courses, when the latter was the only source of training. Since the means in Table 17 indicated some potential differences, t tests were run at grades E-4 and E-5. The comparison made was between correspondence only graduates and the group which had both correspondence and formal school training (Table 19). In grade E-4 formal school training enhanced JKT performance somewhat judging

Table 19

t tests for Grades E-4 and E-5 Correspondence Only vs Both Correspondence and Formal School Training

	t	(Strength of Association)	df
E-4	2.175*	.067	67
E-5	2.767**	.134	49

* $p < .05$

**$p < .01$

from a significant t ratio qualified by a strength of association below 0.10. In grade E-5 the effect was more pronounced with a higher level of significance and a stronger association.

Table 20 presents the average JKT scores for various educational categories. An analysis of variance across these categories was significant. ($F_{(3,352)} = 14.98$, $P < .01$, $Omega^2 = .105$). The relevance of education becomes most pronounced when the entire data set is included and paygrade is ignored. The post testing of differences between the various training groups is indicated in Table 21, which includes as do all similar tables, a schematic description of where significant differences actually lie. The one general conclusion that evolves from Table 21 is

Table 20

Average JKT Scores as a Function of Different Types of Training, Grades Pooled

No School		Correspondence Only		Formal School No Correspondence		Both Formal & Correspondence	
Mean	No.	Mean	No.	Mean	No.	Mean	No.
36.9	33	46.07	71	43.23	82	51.42	170

Table 21

Differences Between Average JKT Scores on the Basis of Different Training Categories for Pooled Paygrades (The Newman-Keuls Procedure)

		No School	Formal School Only	Correspondence Only	Both Formal & Correspond.
	Means	36.9	43.33*	46.07*	51.42*
No School	36.9		6.33*	9.17*	14.52*
Formal School Only	43.23			2.84	8.19*
Corresp. Only	46.07				5.35
Both Formal & Corresp.	51.42				

* $p < .05$

Schematic Presentation of Significant Differences

No School	Formal School	Correspondence	Formal School & Correspondence

Categories which share common underlining are not significantly different.

that training of any kind is better than none. However, this result is contaminated by the fact that those who have no training are in the lowest paygrades and have the least experience. Further, the more time an individual has in the service, the more experience and school opportunities are presented to him. It could be the combination of these which is the basis of the strong relationship between paygrade (directly related to the time in service) and JKT scores.

Conclusions

1. The JKT was internally reliable and reasonably valid against supervisors ratings.

2. The JKT measured what it attempted to measure - job related information. It was not correlated with "Intelligence Scores" from the Navy entrance test battery.

3. JKT performance increased in direct relation with paygrade. The combination of experience and education may have worked to provide advanced knowledge commensurate with paygrade for Navy cooks.

4. 'A' School graduates performed better on the JKT than non-graduates, in all grade levels tested except E-6.

5. Watch captains did not perform better on the JKT than their subordinates in the same paygrades.

6. Former Stewards tended to score higher than former Commissarymen.

7. When all paygrades were pooled and comparisons were made across types of training including correspondence courses, any training was better than none. This implies a necessity for the continuation of some sort of formal programs.

APPENDIX

Table A-1

Table of Differences Between Means - The results of Analysis of Variance and the Newman - Keuls Procedure.

Stewards - East

ANOVA
F=10.76**
** p<.01
* p<.05

Grade		2	3	4	5	6
	Means	43.33	43.78	54.53	60.29	68.16
2	43.33		.45	11.2	16.96*	24.83*
3	43.78			10.75	16.51*	24.38*
4	54.53				5.76	13.63
5	60.29					7.87
6	68.19					

Grades Sharing common underline do not differ: Schematic Presentation of Differences

<u>234</u>
<u>56</u>

Table A-2

Commissarymen - East

F=14.44**
*P < .05
**P < .01

Grade		3	2	5	4	6
	Means	39.16	40.76	45.78	46.48	60.78
3	39.16		1.6	6.62	7.32	21.62*
2	40.76			5.02	5.72	20.02*
5	45.78				0.7	15.00*
4	46.48					14.3 *
6	60.78					

Grades sharing common underline do not differ: Schematic Presentation of Differences

2345
6

Table A-3

Stewards and Commissarymen Pooled - East

TOTAL:

F=24.756**
*P < .05
**P < .01

Grade		2	3	4	5	6
	Means	38.67	41.40	49.65	54.36	63.00
2	38.67		2.73	10.98*	15.69*	24.33*
3	41.40			8.25*	12.96*	21.6*
4	49.65				4.71	13.35*
5	54.36					8.64*
6	63.00					

Grades sharing common underline do not differ: Schematic Presentation of Differences.

<u>23</u> <u>45</u> <u>6</u>

Table A-4

Stewards - West

ANOVA
F=7.395**
*P < .05
**P < .01

Grade		2	3	4	5	6
	Means	36.66	37.7	54.94	57.75	60.25
2	36.66		1.04	18.28*	21.09*	23.59*
3	37.7			17.24*	20.05*	22.55*
4	54.94				2.81	5.31
5	57.75					2.5
6	60.25					

Grades sharing common underline do not differ: Schematic Presentation of Differences.

23 456

Table A-5

Commissarymen - West

F=8.82**
*P < .05
**P < .01

Grade		2	3	4	5	6
	Means	37.81	38.86	46.14	46.84	60.08
2	37.81		1.05	8.33	9.03	22.27*
3	38.86			7.28	7.98	21.22*
4	46.14				0.7	13.94*
5	46.84					13.24*
6	60.08					

Grades sharing common underline do not differ: Schematic Presentation of Differences.

2345 6

Table A-6

Commissarymen and Stewards Pooled - West

TOTAL:
F=18.90**
*P < .05
**P < .01

Grade		2	3	4	5	6
	Means	37.48	38.22	50.52	52.73	60.14
2	37.48		.74	13.04*	15.25*	22.66*
3	38.22			12.3*	14.51*	21.92*
4	50.52				2.21	9.62*
5	52.73					7.41
6	60.14					

Grades sharing common underline do not differ: Schematic Presentation of Differences.

<u>23</u> <u>4</u> <u>56</u>

Table A-7

Stewards - East and West Pooled

F=16.529**
*P < .05
**P < .01

Grade		2	3	4	5	6
	Means	40.0	41.49	54.78	59.24	63.64
2	40.0		1.49	14.78*	19.24*	23.64*
3	41.49			13.29*	17.75*	22.15*
4	54.78				4.46	8.86
5	59.24					4.40
6	63.64					

Grades sharing common underline do not differ: Schematic Presentation of Differences.

23 456

Table A-8

Commissarymen East and West Pooled

ANOVA:
F=22.22**
*P < .05
**P < .01

Grade		2	3	4	5	6
	Means	38.56	39.02	46.30	46.33	60.46
2	38.56		.46	7.74*	7.77*	21.90*
3	39.02			7.28*	7.31*	21.44*
4	46.30				.03	14.16*
5	46.33					14.13
6	60.46					

Grades sharing common underline do not differ: Schematic Presentation of Differences.

<u>23</u> <u>45</u> <u>6</u>

References

Barrett, R.S. Performance rating. Chicago: Science Research, 1966.

Bureau of Naval Personnel. Commissaryman 3+2 rate training manual (NAVPERS 10279-E). U.S. Government Printing Office Washington; (1971a).

Bureau of Naval Personnel. Steward 3+2 rate training manual (NAVPERS 10694-D). U.S. Government Printing Office: Washington, (1971b).

Foley, J.P. Evaluating maintenance performance: an analysis. (AFHRL-TR-74-57(1)), Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, Oct., 1974.

Gagne, R.M. and Fleishman, E.A. Psychology and human performance. New York: Holt, 1957.

Guilford, J.P. Psychometric methods, McGraw: New York, 1954.

Lawshe, C.H. Principles of personnel testing. New York: McGraw, 1948.

Linton, M. and Gallo, P.S. The Practical statistician: simplified handbook of statistics. Monterey: Brooks, 1975.

Merenda, P.F. The relative effectiveness of formal school and on the job methods of training apprentices in naval occupations. Personnel psychology, 1958, 11, 379-389.

Naval Training Command. Commissaryman 1+C rate training manual (NAVTRA 10280-F). U.S. Government Printing Office: Washington, (1973).

Naval Training Command. Steward 1+C rate training manual (NAVTRA 10695-D). U.S. Government Printing Office: Washington, (1972).

Siegel, A.I., Bergman, B.A., Federman, P., Sellman, W.S., Some techniques for the evaluation of technical training courses and student. (AFHRL-TR-72-15) Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, Feb. 1972.

Thorndike, R.L. Personnel selection. New York: Wiley, 1949.

Tiffin, J. Industrial psychology. New York: Prentice-Hall, 1952.

Vroom, V.H. Work and motivation. New York: Wiley, 1964.